United States Environmental Protection Agency

Predicting Compound Amenability with Liquid Chromatography Mass Spectrometry to Improve Non-targeted Analysis

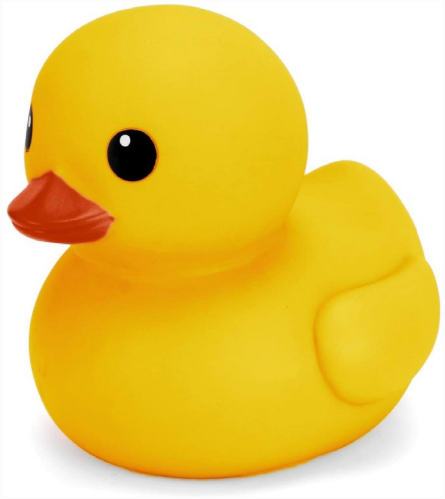
Charles N. Lowe[1], Kristin K. Isaacs[1], Andrew McEachran[2], Christopher M. Grulke[1], Jon R. Sobus[1], Elin M. Ulrich[1], Ann Richard[1], Alex Chao[1], John Wambaugh[1], and Antony J. Williams[1]

1. Center for Computational Toxicology and Exposure, U.S. EPA, Research Triangle Park, NC
2. Agilent Technologies, Inc., Santa Clara, CA

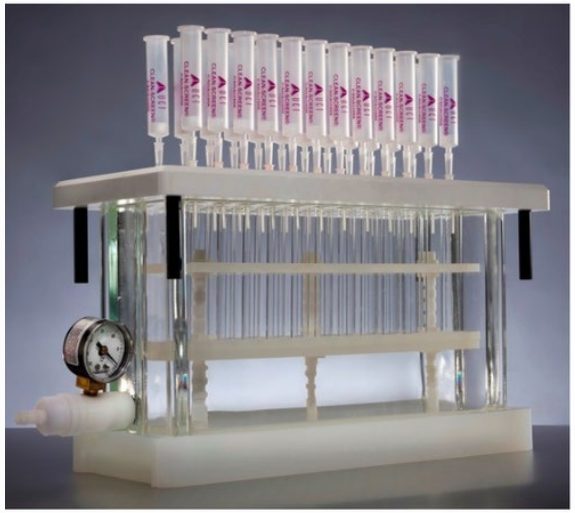
United States Environmental Protection Agency

Disclaimer: The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Complex samples, NTA, and the modeling problem

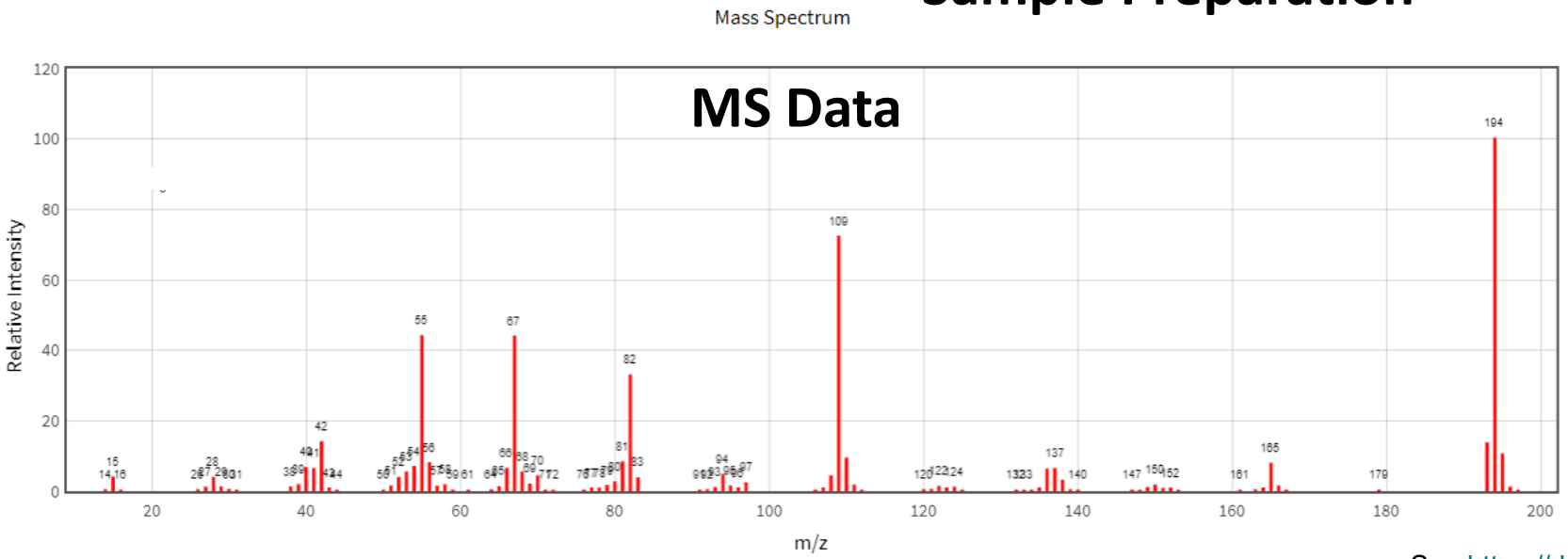


Media Sample



Extraction, Cleanup & Sample Preparation

MS Analysis



???

See https://doi.org/10.1021/acs.est.1c01907 for more on rubber ducky analysis.

Curating a dataset for modeling

MassBank of North America

- **4,103 unique compounds detected in ESI+ LC-MS**
- **3,007 unique compounds detected in ESI- LC-MS**
- **1,542 unique compounds detected in both modes**

- Only amenable compounds identified in MoNA
  - No unamenable compound data
- ToxCast library LC-MS/MS curation
  - Spectra checked individually for quality
    - Provides unamenable compound data

Curating a dataset for modeling

MassBank of North America

- **4,103 unique compounds detected in ESI+ LC-MS**
- **3,007 unique compounds detected in ESI- LC-MS**
- **1,542 unique compounds detected in both modes**

Agilent

- Only amenable compounds identified in MoNA
  - No unamenable compound data
- ToxCast library LC-MS/MS curation
  - Spectra checked individually for quality
    - Provides unamenable compound data

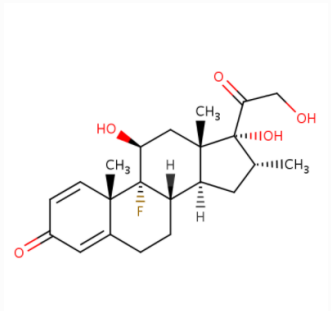
- **ESI+ LC-MS/MS**
  - **393 amenable; 456 unamenable**
- **ESI- LC-MS/MS**
  - **456 amenable; 402 unamenable**

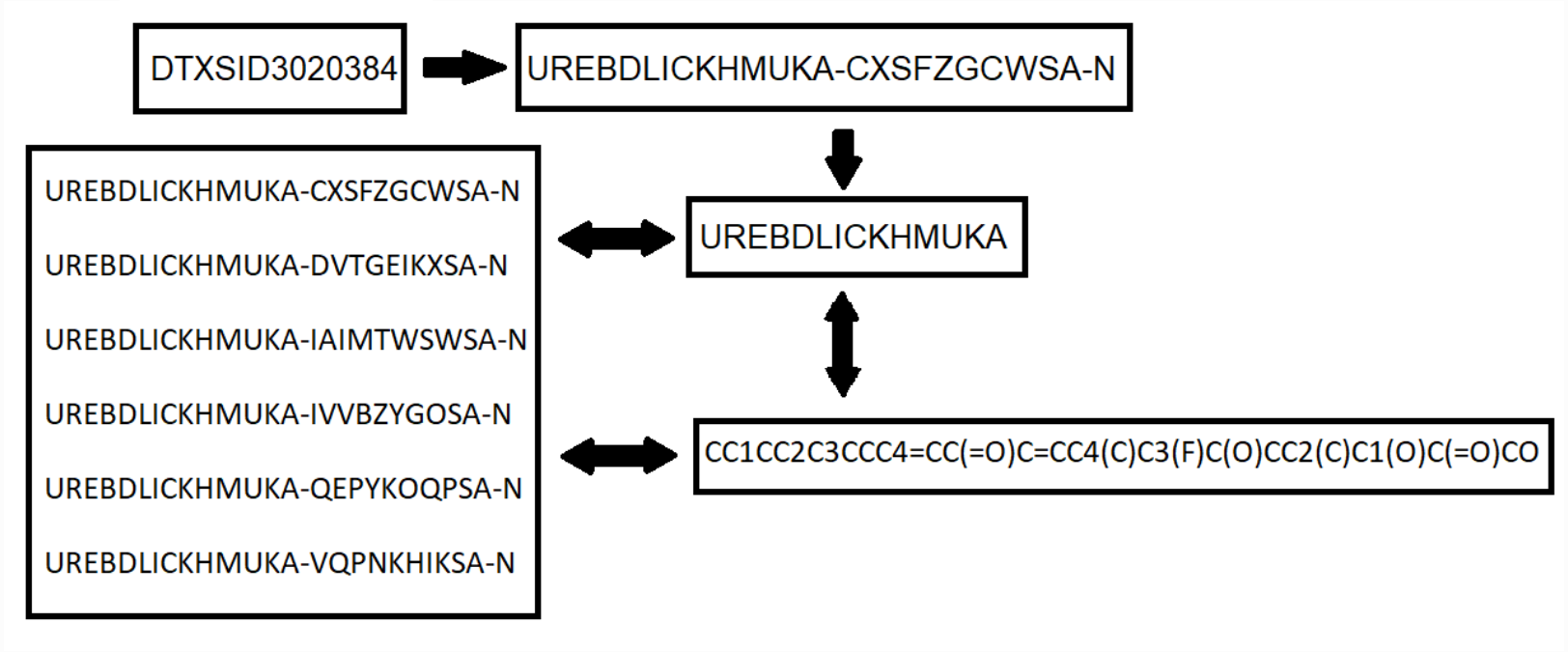
Curating a dataset for modeling

Overall dataset

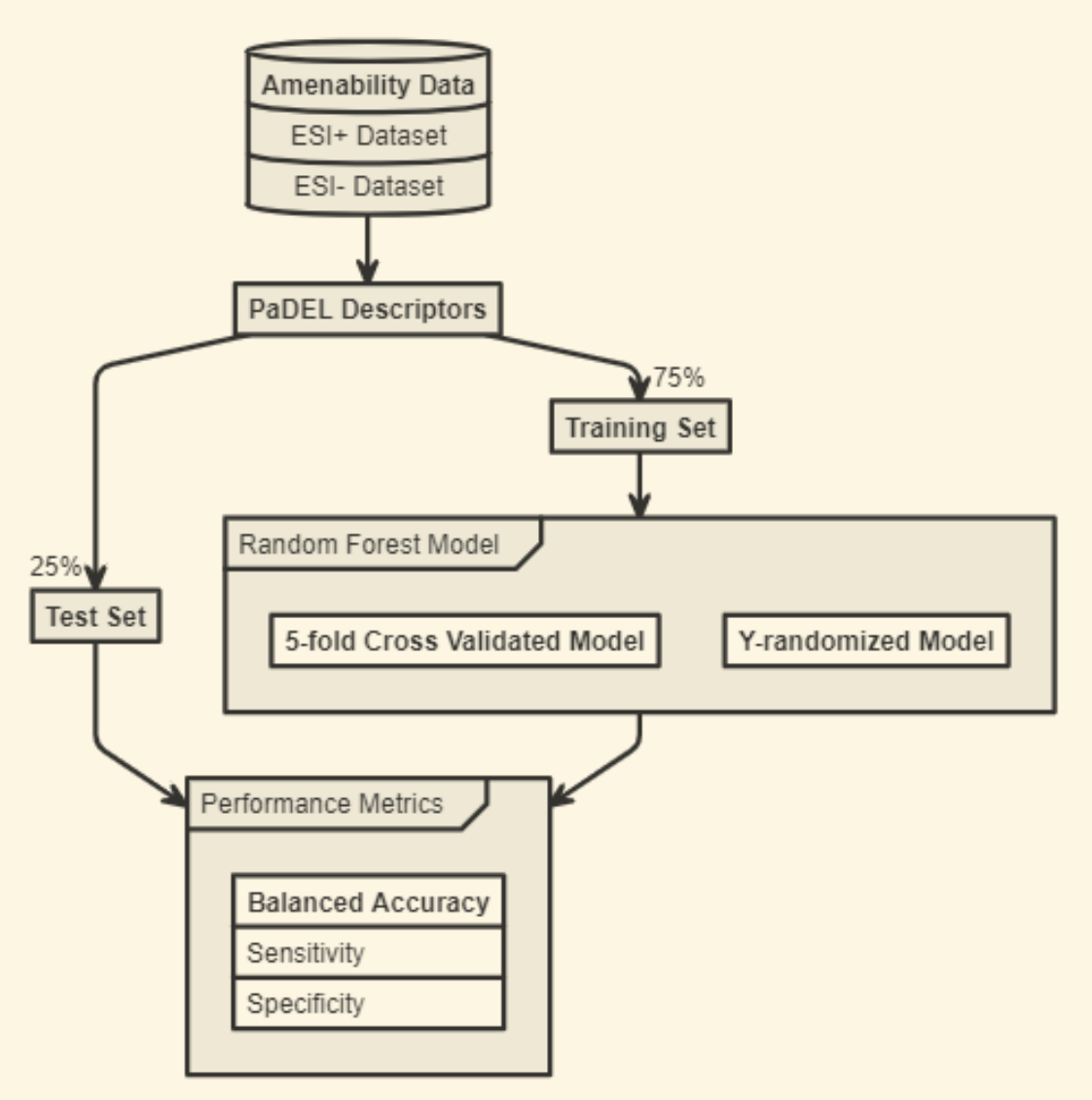
- **ESI+ LC-MS/MS**
  - **4,226 amenable; 387 unamenable**
- **ESI- LC-MS/MS**
  - **3,130 amenable; 360 unamenable**

Curating a dataset for modeling





Machine learning approach



Performance Metrics

$$Sensitivity = \frac{TP}{TP + FN}$$

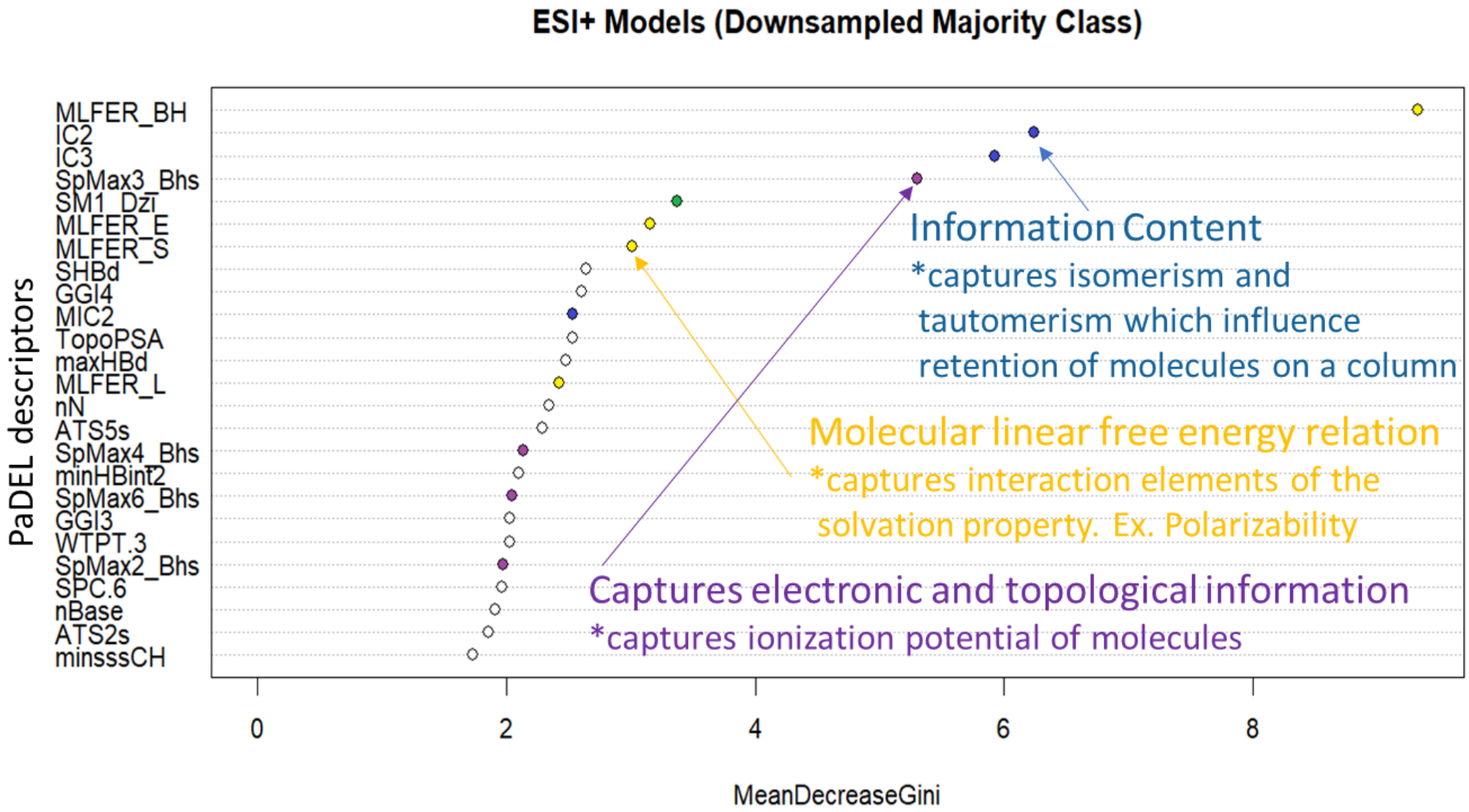
$$Specificity = \frac{TN}{TN + FP}$$

$$\text{balanced accuracy} = \frac{\text{sensitivity} + \text{specificity}}{2}$$

Model performance

	Training Set				Fivefold CV		
Model	Size	Balanced Accuracy	Sensitivity	Specificity	Balanced Accuracy	Sensitivity	Specificity
ESI+ Models (Downsampling Applied)	580	0.78	0.79	0.77	0.77	0.76	0.78
ESI+ Models (Upsampling Applied)	6340	0.99	1.00	0.99	0.99	0.98	1.00
ESI- Models (Downsampling Applied)	550	0.83	0.82	0.84	0.81	0.83	0.79
ESI- Models (Upsampling Applied)	4688	0.99	1.00	0.98	0.98	0.97	1.00
	Test Set				**Y-randomization**		
Model	Size	Balanced Accuracy	Sensitivity	Specificity	Balanced Accuracy	Sensitivity	Specificity
ESI+ Models (Downsampling Applied)	1153	0.81	0.85	0.76	0.48	0.44	0.51
ESI+ Models (Upsampling Applied)	1153	0.58	0.98	0.19	0.55	0.48	0.63
ESI- Models (Downsampling Applied)	871	0.82	0.85	0.80	0.50	0.49	0.51
ESI- Models (Upsampling Applied)	871	0.68	0.99	0.38	0.51	0.46	0.56

Mechanistic interpretation

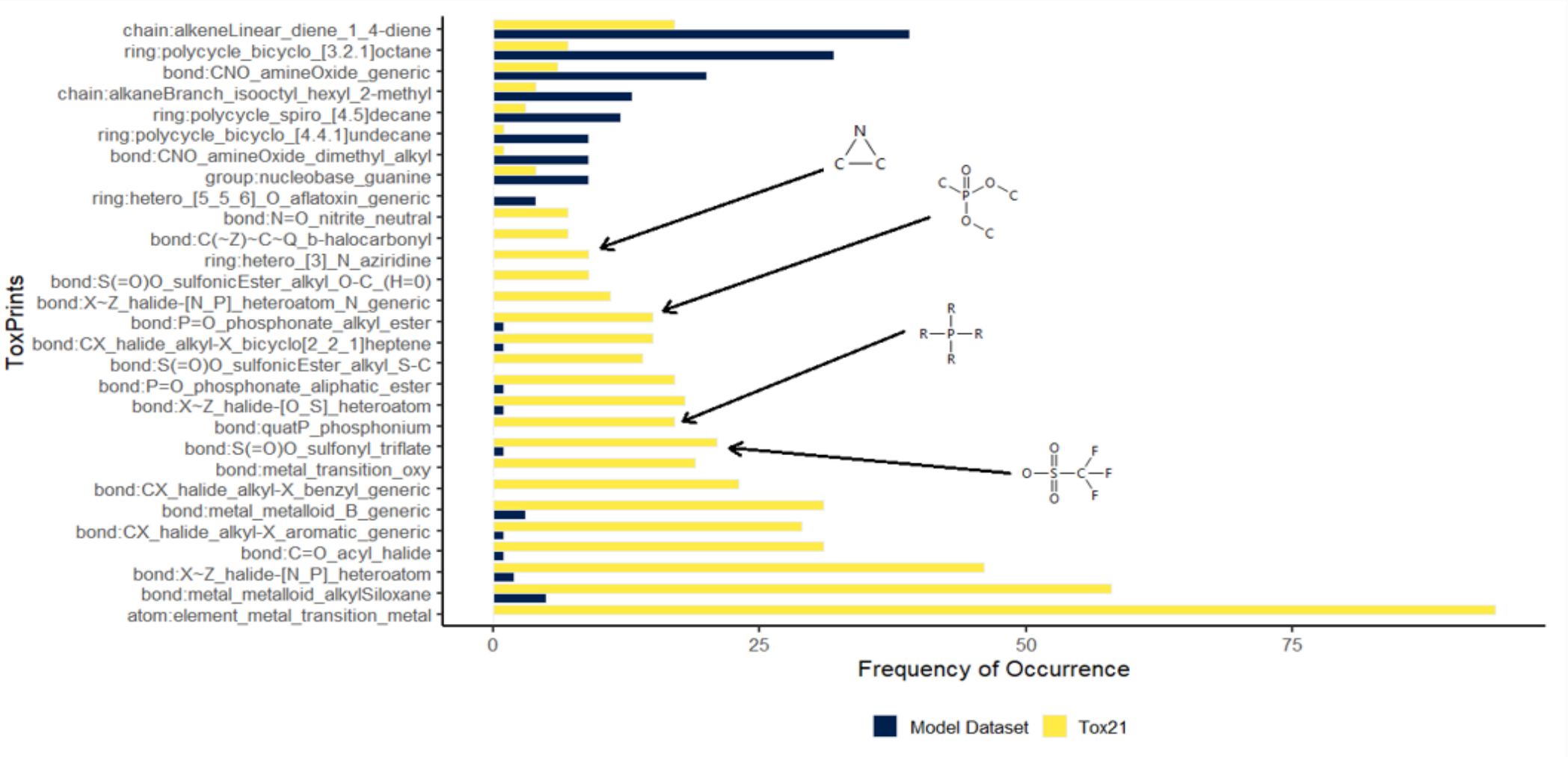


External validation

- 1,768 chemical compounds analyzed for LC-MS amenability as part of ToxCast program
  - All compounds exclusive of modeling dataset

ESI- Downsampled Model		
	Amenable (Prediction)	Unamenable (Prediction)
Detected (Experiment)	323	502
Not-detected (Experiment)	68	874
Sensitivity	**0.83**	
Specificity	**0.64**	
Balanced Accuracy	**0.73**	
ESI+ Downsampled Model		
	Amenable (Prediction)	Unamenable (Prediction)
Detected (Experiment)	423	402
Not-detected (Experiment)	103	839
Sensitivity	**0.80**	
Specificity	**0.68**	
Balanced Accuracy	**0.74**	
Combined Models		
	Amenable (Prediction)	Unamenable (Prediction)
Detected (Experiment)	505	320
Not-detected (Experiment)	129	813
Sensitivity	**0.80**	
Specificity	**0.72**	
Balanced Accuracy	**0.76**	

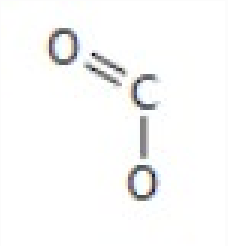
Model applicability to ToxCast



Comparison of prevalent ToxPrint chemotypes in amenability dataset against the ToxCast dataset

Model comparison with expert intuition

- A small molecule containing a carboxylic acid functional group *should* be amenable to ESI- LC-MS

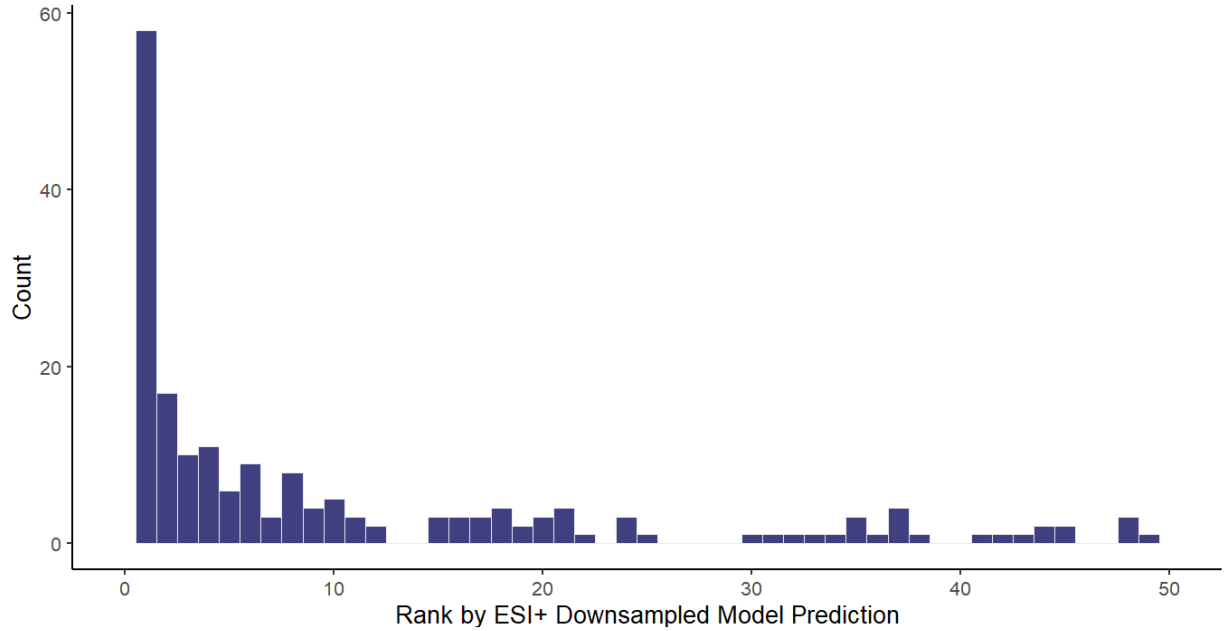


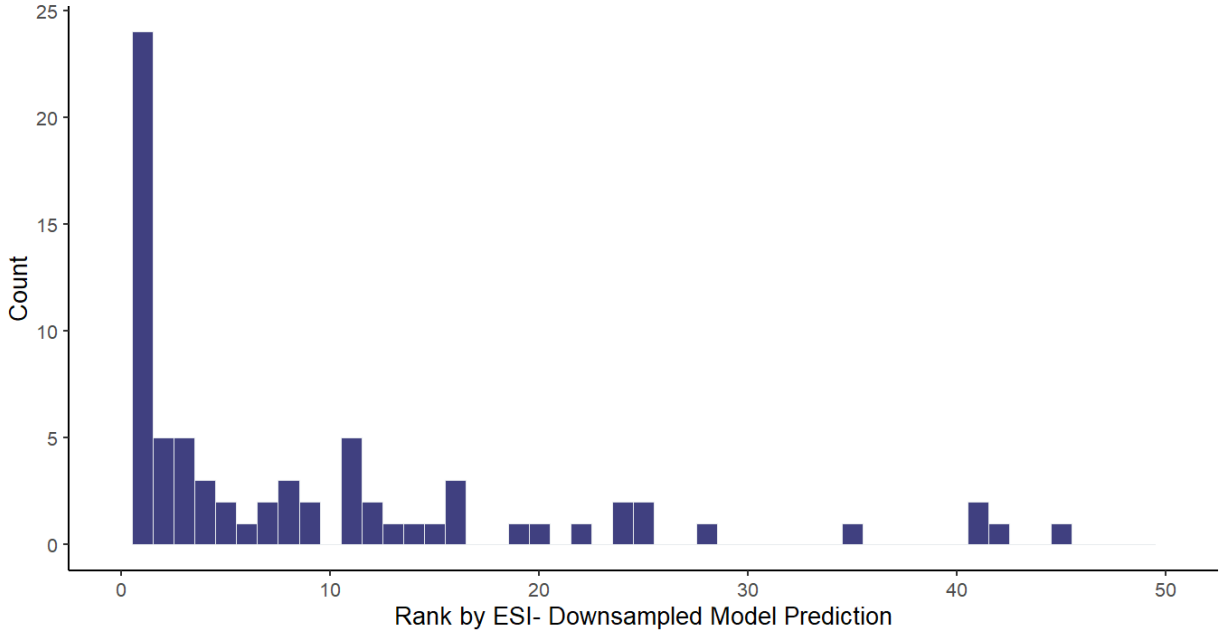
- 773 compounds contained the ToxPrint “bond:C(=O)O_carboxylicAcid_generic” in amenability dataset

ESI- Downsampled Model		
	Amenable (Prediction)	Unamenable (Prediction)
Detected (Experiment)	728	4
Not-detected (Experiment)	37	9
Sensitivity	**0.95**	
Specificity	**0.69**	
Balanced Accuracy	**0.82**	

Suspect-screening application

- List of ENTACT compounds identified in ESI+ & ESI- LC-MS
  - 228 in ESI+
  - 108 in ESI-
- Retrieved candidates for each molecular formula via Dashboard
  - 13,325 candidates for ESI+
  - 7,079 candidates for ESI-
- Generated amenability predictions for candidate structures
- Rank ordered candidates by amenability probability (probability of detection)





Summary

- Amenability models are available as a deployable API inside of a Docker container
- These models will be used in the ChemSpace tool currently in early development
  - Please check out our ChemSpace poster!

Research Paper | Published: 14 October 2021

Predicting compound amenability with liquid chromatography-mass spectrometry to improve non-targeted analysis

Charles N. Lowe, Kristin K. Isaacs, Andrew McEachran, Christopher M. Grulke, Jon R. Sobus, Elin M. Ulrich, Ann Richard, Alex Chao, John Wambaugh & Antony J. Williams

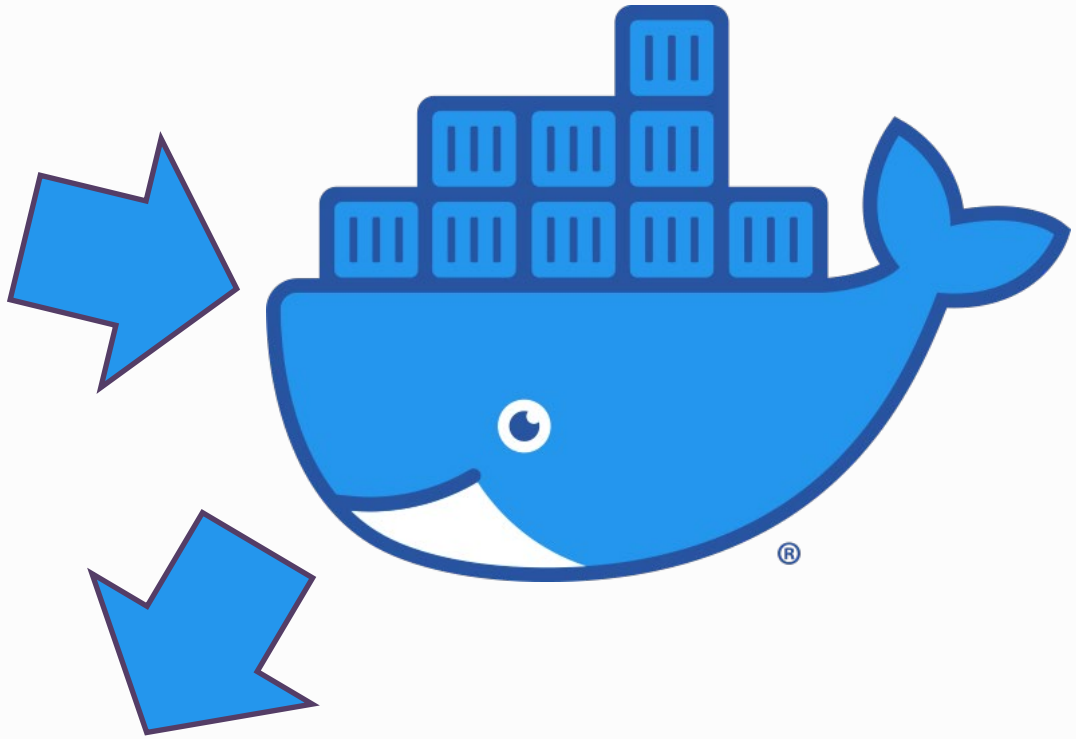
Analytical and Bioanalytical Chemistry **413**, 7495–7508 (2021) | Cite this article

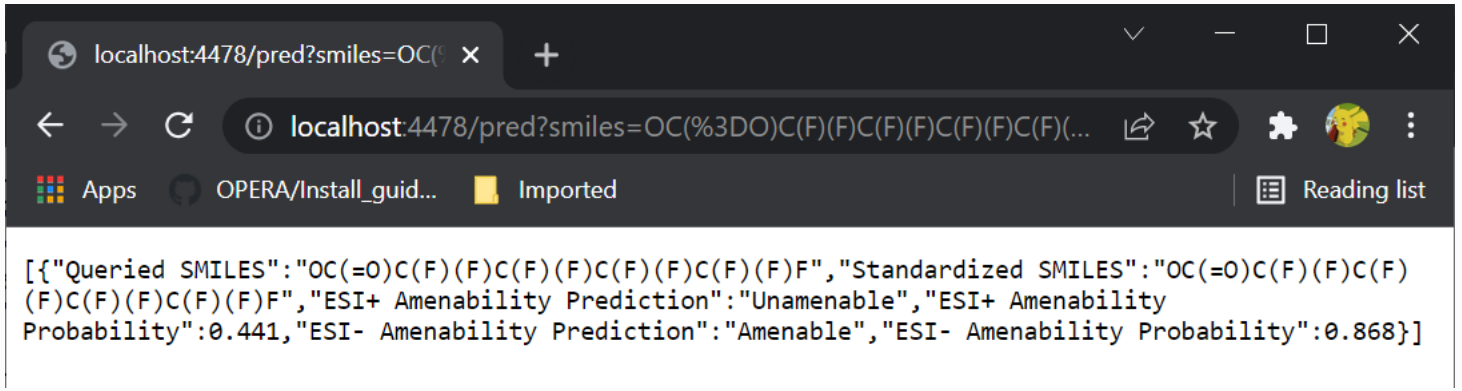
Research Paper | Published: 14 October 2021

Predicting compound amenability with liquid chromatography-mass spectrometry to improve non-targeted analysis

Charles N. Lowe, Kristin K. Isaacs, Andrew McEachran, Christopher M. Grulke, Jon R. Sobus, Elin M. Ulrich, Ann Richard, Alex Chao, John Wambaugh & Antony J. Williams

Analytical and Bioanalytical Chemistry **413**, 7495–7508 (2021) | Cite this article





Contributing researchers

EPA ORD

Hussein Al-Ghoul*
Alex Chao
Louis Groff
Jarod Grossman*
Kristin Isaacs
Sarah Laughlin*
Hannah Liberatore
James McCord
Kelsey Miller
Jeff Minucci
Seth Newton
Katherine Phillips
Allison Phillips*
Tom Purucker
Randolph Singh*
Jon Sobus
Mark Strynar
Elin Ulrich
Nelson Yeung*

EPA ORD (cont.)

Kathie Dionisio
Chris Grulke
Kamel Mansouri*
Andrew McEachran*
Ann Richard
Adam Swank
John Wambaugh
Antony Williams

Agilent

Jarod Grossman
Andrew McEachran

GDIT

Ilya Balabin
Tom Transue
Tommy Cathey

* = ORISE/ORAU


Thank you for
Listening!